

# THE UNITED STATES OF AMERICA

**TO ALL TO WHOM THESE PRESENTS SHALL COME:**

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office

May 27, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/502,995

FILING DATE: *September 15, 2003*

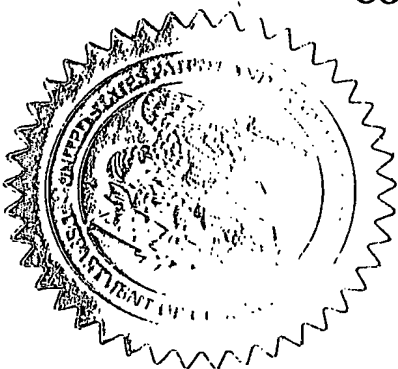
RELATED PCT APPLICATION NUMBER: PCT/US04/07451

REC'D 01 JUN 2004

WIPO

PCT

By Authority of the  
COMMISSIONER OF PATENTS AND TRADEMARKS



*M. Tarver*

M. TARVER  
Certifying Officer

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**PROVISIONAL APPLICATION FOR PATENT COVER SHEET**

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(S)					
Given Name (first and middle (if any))		Family Name or Surname		Residence (City and either State or Foreign Country)	
Barbara K. Dawn C. J. Raymond L.		Zehentner-Wilkinson Hayes Houghton		Poulsbo, Washington Preston, Washington Bothell, Washington	
<input type="checkbox"/> Additional inventors are being named on the _____ separately numbered sheets attached hereto					
TITLE OF THE INVENTION (500 characters max)					
METHODS, COMPOSITIONS AND KITS FOR THE DETECTION AND MONITORING OF LUNG CANCER					
CORRESPONDENCE ADDRESS					
Direct all correspondence to:					
<input checked="" type="checkbox"/> Customer Number		32111		32111	
OR		Type Customer Number here			
<input type="checkbox"/> Firm or Individual Name					
Address					
Address					
City		State		ZIP	
Country		Telephone		Fax	
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification Number of Pages		35		<input type="checkbox"/> CD(s), Number _____	
<input type="checkbox"/> Drawing(s) Number of Sheets		_____		<input checked="" type="checkbox"/> Other (specify) _____	
<input type="checkbox"/> Application Data Sheet. See 37 CFR 1.76		Sequence Listing (16 pages); Fee Transmittal (+ copy); Postcard.			
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT					
<input checked="" type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.					
<input type="checkbox"/> A check or money order for \$_____ is enclosed to cover the filing fees.					
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees to Deposit Account Number: _____					
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any deficiency or credit any overpayment to Deposit Account Number: _____					
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.					
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.					
<input checked="" type="checkbox"/> No.					
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: _____					

Respectfully submitted,

SIGNATURE

TYPED or  
PRINTED NAME

TELEPHONE

Susan E. Lingenfelter

206-754-5898

DATE

REGISTRATION NO.  
(if appropriate)

DOCKET NUMBER:

September 15, 2003

41,156

609P2

**USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT**

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Commissioner for Patents, Washington, D.C. 20231.

F:\Pat\PTOSB16.doc

22388 U.S. PTO  
60/502995

**FEE TRANSMITTAL  
for FY 2003**

Patent fees are subject to annual revision.

☒ Applicant claims small entity status. See 37 CFR 1.27.**TOTAL AMOUNT OF PAYMENT (\$)** **80.00**

Complete if Known

Filing Date	September 15, 2003
First Named Inventor	Barbara K. Zehentner-Wilkinson
Examiner Name	
Group Art Unit	
Attorney Docket No.	609P2

**METHOD OF PAYMENT**

☐ Payment Enclosed:  
☐ Check ☐ Credit card ☐ Money Order ☐ Other

☒ Deposit Account:  
 Deposit Account Number: 50-0597  
 Deposit Account Name: Corixa Corporation

The Commissioner is authorized to (check all that apply)

☒ Charge fee(s) indicated below ☒ Credit any overpayments  
☐ Charge any additional fee(s) during the pendency of this application  
☐ Charge fee(s) indicated below, except for the filing fee  
☒ Charge any deficiencies  
 to the above-identified deposit account.

**FEE CALCULATION****1. BASIC FILING FEE**

Large Entity		Small Entity		Fee Description	Fee Paid
Fee Code	Fee (\$)	Fee Code	Fee (\$)		
1001	750	2001	375	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	
<b>SUBTOTAL (1)</b>					<b>(5) 80</b>

**2. EXTRA CLAIM FEES**

Total Claims	Extra Claims	Fee from below	Fee Paid
	** =		
Independent Claims	** =		
Multiple Dependent	*		

Large Entity		Small Entity		Fee Description
Fee Code	Fee (\$)	Fee Code	Fee (\$)	
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent
<b>SUBTOTAL (2)</b> <b>(5)</b>				

\*\*or number previously paid, if greater; For Reissues, see above

**FEE CALCULATION (continued)****3. ADDITIONAL FEES**

Large Entity		Small		Fee Description	Fee Paid
Fee Code	Fee (\$)	Fee Code	Fee (\$)		
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2520	1812	2520	For filing a request for <i>ex parte</i> reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1840*	1805	1840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1450	2254	725	Extension for reply within fourth month	
1255	1970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1510	1451	1510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1300	2453	650	Petition to revive - unintentional	
1501	1300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee for provisional applications	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR § 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR § 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) \_\_\_\_\_

\*Reduced by Basic Filing Fee Paid

**SUBTOTAL (3)** **(5)**

Name (Print/Type)	Susan E. Lingenfelter	Registration No. Attorney/Agent	41,156
Firm Name/Address	Corixa Corporation, 1124 Columbia Street, Suite 200, Seattle, WA 98104		
Signature	<i>Susan E. Lingenfelter</i>	Date	September 15, 2003

**32111**

PATENT TRADEMARK OFFICE

METHODS, COMPOSITIONS AND KITS  
FOR THE DETECTION AND MONITORING OF LUNG CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to the field of cancer diagnostics.

- 5 More specifically, the present invention relates to methods, compositions and kits for the detection of lung cancer in patients with different type, stage and grade of tumors that employ oligonucleotide hybridization and/or amplification to simultaneously detect two or more tissue-specific polynucleotides in a biological sample suspected of containing lung cancer cells.

10 BACKGROUND OF THE INVENTION

Field of the Invention

- Lung cancer remains a significant health problem throughout the world. The failure of conventional lung cancer treatment regimens can commonly be attributed, in part, to delayed disease diagnosis. Although significant advances have been made in the area of  
15 lung cancer diagnosis, there still remains a need for improved detection methodologies that permit early, reliable and sensitive determination of the presence of lung cancer cells.

Description of the Related Art

- Lung cancer has the highest mortality rate of any of the cancers and is one of the  
20 most difficult to diagnose early. There are an estimated 1 million deaths annually worldwide for this disease. According to the American Cancer Society in 2002 alone there were an estimated 169,200 new cases diagnosed and ~ 154,900 deaths. Typically lung cancers are classified into two major types: Non-Small Cell Lung Carcinomas (NSCLC) comprising squamous, adeno and large cell carcinomas and Small Cell Lung Carcinomas  
25 (SCLC). These groups represent ~75% and 25% of all lung tumors respectively with adenocarcinoma and squamous cell carcinoma being the most prevalent forms of NSCLC with large cell carcinomas being ~10%. Within the group of NSCLC, adenocarcinoma is

currently the most prominent form of lung cancer in younger persons, women of all ages, lifetime nonsmokers and long-term former smokers. SCLC typically fall into two subtypes oat cell and intermediate cell. Less common tumors include carcinoid and mesotheliomas among others but these represent only a small percentage of all lung tumors. In almost all cases early diagnosis of NSCLC is elusive and most lung cancers have already metastasized by the time they are detected. Only 16.7% are localized on initial diagnosis. If tumors can be detected at a point where they are confined then the combination of chemotherapy and radiation has a possibility of success but overall the 5year prognosis is very poor with only 10-15% survival rate. The picture with SCLC is even bleaker only 6% localized at initial diagnosis and with 5 year survival rates of ~6%.

X-ray and computer tomography of the chest and abdomen are frequently used in diagnosis of lung tumors but lack sensitivity for detecting small foci and usually detect tumors that have already metastasized. Sputum cytology as a potential screening method in high-risk individuals has only been partially effective and often does not yield tumor type. To stage the disease CAT scan, MRI or bone scans are used to evaluate the spread of disease. Treatment for lung cancer is typically surgical, radiological or chemotherapy or combinations thereof, but usually with poor outcome due to the late diagnosis of disease.

The current tests for lung cancer lack either the clinical sensitivity to detect early tumors or provide inadequate stage/grade information or lack tumor specificity due to their originating from other tumor types or being present in benign lung disorders. There is therefore a need to develop specific tests that can improve lung cancer diagnosis and prognosis and potentially differentiate between NSCLC and SCLC. The present invention achieves these and other related objectives by providing methods that are useful for the identification of tissue-specific polynucleotides, in particular tumor-specific polynucleotides, as well as antibodies and methods, compositions and kits for the detection and monitoring of cancer cells in a patient afflicted with the disease.

## SUMMARY OF THE INVENTION

The present invention provides methods for detecting the presence of lung cancer cells in a patient. Such methods comprise the steps of: (a) obtaining a biological sample from the patient; (b) contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof (c) amplifying the polynucleotides; and (d) detecting the amplified polynucleotides; wherein the presence of one or more of the amplified polynucleotides indicates the presence of lung cancer cells in the patient.

By some embodiments, detection of the amplified polynucleotides may be preceded by a fractionation step such as, for example, gel electrophoresis. Alternatively or additionally, detection of the amplified polynucleotides may be achieved by hybridization of a labeled oligonucleotide probe that hybridizes specifically, under moderately stringent conditions, to such polynucleotides. Oligonucleotide labeling may be achieved by incorporating a radiolabeled nucleotide or by incorporating a fluorescent label.

In certain preferred embodiments, cells of a specific tissue type may be enriched from the biological sample prior to the steps of detection. Enrichment may be achieved by a methodology selected from the group consisting of cell capture and cell depletion. Exemplary cell capture methods include immunocapture and comprise the steps of: (a) adsorbing an antibody to a tissue-specific cell surface to cells said biological sample; (b) separating the antibody adsorbed tissue-specific cells from the remainder of the biological sample. Exemplary cell depletion may be achieved by cross-linking red cells and white cells followed by a subsequent fractionation step to remove the cross-linked cells.

Alternative embodiments of the present invention provide methods for determining the presence or absence of lung cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins; (b) detecting in the sample a level of at least one of the polynucleotides (such as, for example,

mRNA) that hybridize to the oligonucleotides; and (c) comparing the level of polynucleotides that hybridize to the oligonucleotides with a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction  
5 using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

10 In related aspects, methods are provided for monitoring the progression of lung cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode lung tumor proteins; (b) detecting in the sample an amount of the polynucleotides that hybridize to the oligonucleotides; (c) repeating steps (a) and (b)  
15 using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Certain embodiments of the present invention provide that the step of amplifying said first polynucleotide and said second polynucleotide is achieved by the polymerase  
20 chain reaction (PCR).

The present invention also provides kits that are suitable for performing the detection methods of the present invention. Exemplary kits comprise oligonucleotide primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Within certain embodiments, kits according to the present invention may also comprise a nucleic  
25 acid polymerase and suitable buffer.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

**BRIEF DESCRIPTION OF SEQUENCE IDENTIFIERS**

- SEQ ID NO: 1 is the determined cDNA sequence L762P.
- SEQ ID NO: 2 is the amino acid sequence encoded by the sequence of SEQ ID NO: 1.
- 5 SEQ ID NO: 3 is the determined cDNA sequence L984P.
- SEQ ID NO: 4 is the amino acid sequence encoded by the sequence of SEQ ID NO: 3.
- SEQ ID NO: 5 is the determined cDNA sequence L550S.
- SEQ ID NO: 6 is the amino acid sequence encoded by the sequence of SEQ ID NO: 5.
- SEQ ID NO: 7 is the determined cDNA sequence L552S.
- 10 SEQ ID NO: 8 is the amino acid sequence encoded by the sequence of SEQ ID NO: 7.
- SEQ ID NO: 9 is the DNA sequence of L552S INT forward primer.
- SEQ ID NO: 10 is the DNA sequence of L552S INT reverse primer.
- SEQ ID NO: 11 is the DNA sequence of L552S Taqman probe.
- SEQ ID NO: 12 is the DNA sequence of L550S INT forward primer.
- 15 SEQ ID NO: 13 is the DNA sequence of L550S INT reverse primer.
- SEQ ID NO: 14 is the DNA sequence of L550S Taqman probe.
- SEQ ID NO: 15 is the DNA sequence of L726P INT forward primer.
- SEQ ID NO: 16 is the DNA sequence of L726P INT reverse primer.
- SEQ ID NO: 17 is the DNA sequence of L726P Taqman probe.
- 20 SEQ ID NO: 18 is the DNA sequence of L984P INT forward primer.
- SEQ ID NO: 19 is the DNA sequence of L984P INT reverse primer.
- SEQ ID NO: 20 is the DNA sequence of L984P Taqman probe.
- SEQ ID NO: 21 is the determined cDNA sequence of L763P.
- SEQ ID NO: 22 is the amino acid sequence encoded by the sequence of SEQ ID NO: 21.
- 25 SEQ ID NO: 23 is the DNA sequence of L763P INT forward primer.
- SEQ ID NO: 24 is the DNA sequence of L763P reverse primer.
- SEQ ID NO: 25 is the DNA sequence of L763P Taqman probe.
- SEQ ID NO: 26 is the determined cDNA sequence of L587.
- SEQ ID NO: 27 is the amino acid sequence encoded by the sequence of SEQ ID NO: 26.
- 30 SEQ ID NO: 28 is the DNA sequence of L587 INT forward primer.



- SEQ ID NO:29 is the DNA sequence of L587 INT reverse primer.  
 SEQ ID NO:30 is the DNA sequence of L587 Taqman probe.  
 SEQ ID NO:31 is the determined cDNA sequence of L523.  
 SEQ ID NO:32 is the amino acid sequence encoded by the sequence of SEQ ID NO:31.  
 5 SEQ ID NO:33 is the DNA sequence of L523 primer.  
 SEQ ID NO:34 is the DNA sequence of L523 primer.

#### DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is directed generally to methods that are suitable for the identification of tissue-specific polynucleotides as well as to methods,  
 10 compositions and kits that are suitable for the diagnosis and monitoring of lung cancer, in particular such methods, compositions and kits are suitable for use in the diagnosis, differentiation and/or prognosis of NSCLC and SCLC. Such diagnostic methods will form the basis for a molecular diagnostic test for detecting lung cancer metastases in lung tissue and for the detection of anchorage independent lung cancer cells in blood as well as in  
 15 mediastinal lymph nodes of distant metastases.

A variety of genes have been identified as over-expressed in lung tumors, in particular squamous or adeno forms of NSCLC or small cell carcinomas. These include, but are not limited to: L762P, L984P, L550S/L548S, L552S/L547S, L552/L547S, L200T, L514S, L551S, L587S, L763S, L773P, L801P, L985P, L1058C, L1081C, L523S,  
 20 OF1783P, B307D (WIPO International Patent Application Nos: WO 99/47674, published September 23, 1999; WO 00/61612, published October 19, 2000; WO 02/00174, published January 3, 2002; WO 02/47534, published June 20, 2002; WO 01/72295, published October 4, 2001; WO 02/092001, published November 21, 2002; WO 01/00828, published January 1, 2001; WO 02/04514, published January 17, 2002; WO 01/92525, published  
 25 December 6, 2002; WO 02/02623, published January 10, 2002. US Patent Nos: Wang et al., 6,482,597, issued November 22, 2002; Wang et al., 6,518,256, issued February 11, 2003; Wang et al., 6,426,072, issued July 30, 2002; Reed et al., 6,210,883, issued April 3, 2001; Wang et al., 6,504,010, issued January 7, 2003; Wang et al., 6,509,448, issued

January 21, 2003. Wang *et al*; *Oncogene*; 21(49):7598-604, 2002 (collagen type XI alpha 1).).

- These genes were identified and characterized using PCR and cDNA library subtractions as well as electronic subtractions with each of the tumor types individually.
- 5 The cDNAs identified were then evaluated by microarray then by Real Time PCR on tissue panels to identify specific expression patterns. Table 1 highlights the specificity of these genes for either adeno or squamous forms of NSCLC or both as well as genes specific for small cell lung carcinomas. In some cases reactivity with large cell carcinomas has also been identified by Real Time PCR analysis.

10

Table 1

Gene	Squamous	Adeno	Small cell	Large cell	Normal Lung
L762P	+++++	+			
L984P		+	+++		-
L550S/L548S		+++++	+		-
L552S/L547S	++	+++++			-
L200T	+	++			-
L514S	++++	++++		++	-
L551S		++++			-
L587S	+	+	+++	+/-	-
L763P	+++++			+	-
L773P	+++	+++			-
L801P	++++	++++			-
L978P	+	++		++	-
L985P		+	+++++	+/-	-
L1058C			++		-
L1081C			++		-
L523S	+++++	+++++	+	++	-
OF 1783P			+++++		-
B307D	++	++		+	-

#### Identification of Tissue-specific Polynucleotides

- Certain embodiments of the present invention provide methods, compositions and
- 15 kits for the detection of lung cancer cells within a biological sample from patients with different type, stage and grade of tumors. These methods comprise the step of detecting

one or more tissue-specific polynucleotide(s) from a patient's biological sample the over-expression of which polynucleotides indicates the presence of lung cancer cells within the patient's biological sample. Accordingly, the present invention also provides methods that are suitable for the identification of tissue-specific polynucleotides. As used herein, the  
5 phrases "tissue-specific polynucleotides" or "tumor-specific polynucleotides" are meant to include all polynucleotides that are at least two-fold over-expressed as compared to one or more control tissues. As discussed in further detail herein below, over-expression of a given polynucleotide may be assessed, for example, by microarray and/or quantitative real-time polymerase chain reaction (Real-time PCR<sup>TM</sup>) methodologies.

10 Exemplary methods for detecting tissue-specific polynucleotides may comprise the steps of: (a) performing a genetic subtraction to identify a pool of polynucleotides from a tissue of interest; (b) performing a DNA microarray analysis to identify a first subset of said pool of polynucleotides of interest wherein each member polynucleotide of said first subset is at least two-fold over-expressed in said tissue of interest as compared to a control  
15 tissue; and (c) performing a quantitative polymerase chain reaction analysis on polynucleotides within said first subset to identify a second subset of polynucleotides that are at least two-fold over-expressed as compared to said control tissue.

#### Polynucleotides Generally

As used herein, the term "polynucleotide" refers generally to either DNA or RNA  
20 molecules. Polynucleotides may be naturally occurring as normally found in a biological sample such as blood, serum, lymph node, bone marrow, sputum, urine and tumor biopsy samples. Alternatively, polynucleotides may be derived synthetically by, for example, a nucleic acid polymerization reaction. As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may  
25 be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present

invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.* an endogenous sequence that encodes a tumor protein, such as a lung tumor protein, or a portion thereof) or may  
 5 comprise a variant, or a biological or antigenic functional equivalent of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below. The term "variants" also encompasses homologous genes of xenogenic origin.

When comparing polynucleotide or polypeptide sequences, two sequences are said  
 10 to be "identical" if the sequence of nucleotides or amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A  
 15 "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc.,  
 20 Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified  
 25 Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of*

*Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the  
 5 identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

10 One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity  
 15 for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid  
 20 sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the  
 25 sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Therefore, the present invention encompasses polynucleotide and polypeptide sequences having substantial identity to the sequences disclosed herein, for example those comprising at least 50% sequence identity, preferably at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide or polypeptide sequence of this invention using the methods described herein, (*e.g.*, BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

In additional embodiments, the present invention provides isolated polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50,

51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as  
5 promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative DNA  
10 segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

In other embodiments, the present invention is directed to polynucleotides that are  
15 capable of hybridizing under moderately stringent conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a  
20 solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS.

Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a  
25 polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of

one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

5 Microarray Analyses

Polynucleotides that are suitable for detection according to the methods of the present invention may be identified, as described in more detail below, by screening a microarray of cDNAs for tissue and/or tumor-associated expression (*e.g.*, expression that is at least two-fold greater in a tumor than in normal tissue, as determined using a  
10 representative assay provided herein). Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10614-10619 (1996) and Heller *et al.*, *Proc. Natl. Acad. Sci. USA* 94:2150-2155 (1997)).

Microarray is an effective method for evaluating large numbers of genes but due to  
15 its limited sensitivity it may not accurately determine the absolute tissue distribution of low abundance genes or may underestimate the degree of overexpression of more abundant genes due to signal saturation. For those genes showing overexpression by microarray expression profiling, further analysis was performed using quantitative RT-PCR based on Taqman™ probe detection, which comprises a greater dynamic range of sensitivity. Several  
20 different panels of normal and tumor tissues, distant metastases and cell lines were used for this purpose.

Quantitative Real-time Polymerase Chain Reaction

Suitable polynucleotides according to the present invention may be further characterized or, alternatively, originally identified by employing a quantitative PCR  
25 methodology such as, for example, the Real-time PCR methodology. By this methodology, tissue and/or tumor samples, such as, *e.g.*, metastatic tumor samples, may be tested along side the corresponding normal tissue sample and/or a panel of unrelated normal tissue samples.



Real-time PCR (*see* Gibson et al., *Genome Research* 6:995-1001, 1996; Heid et al., *Genome Research* 6:986-994, 1996) is a technique that evaluates the level of PCR product accumulation during amplification. This technique permits quantitative evaluation of mRNA levels in multiple samples. Briefly, mRNA is extracted from tumor and normal  
5 tissue and cDNA is prepared using standard techniques.

Real-time PCR may, for example, be performed either on the ABI 7700 Prism or on a GeneAmp® 5700 sequence detection system (Applied Biosystems, Foster City, CA). The 7700 system uses a forward and a reverse primer in combination with a specific probe with a 5' fluorescent reporter dye at one end and a 3' quencher dye at the other end  
10 (Taqman™). When the Real-time PCR is performed using Taq DNA polymerase with 5'-3' nuclease activity, the probe is cleaved and begins to fluoresce allowing the reaction to be monitored by the increase in fluorescence (Real-time). The 5700 system uses SYBR® green, a fluorescent dye, that only binds to double stranded DNA, and the same forward and reverse primers as the 7700 instrument. Matching primers and fluorescent probes may  
15 be designed according to the primer express program (Applied Biosystems, Foster City, CA). Optimal concentrations of primers and probes are initially determined by those of ordinary skill in the art. Control (*e.g.*,  $\beta$ -actin) primers and probes may be obtained commercially from, for example, Perkin Elmer/Applied Biosystems (Foster City, CA).

To quantitate the amount of specific RNA in a sample, a standard curve is  
20 generated using a plasmid containing the gene of interest. Standard curves are generated using the Ct values determined in the real-time PCR, which are related to the initial cDNA concentration used in the assay. Standard dilutions ranging from  $10^{-10}$  to  $10^{-6}$  copies of the gene of interest are generally sufficient. In addition, a standard curve is generated for the control sequence. This permits standardization of initial RNA content of a tissue sample to the  
25 amount of control for comparison purposes.

In accordance with the above, and as described further below, the present invention provides the illustrative lung tissue- and/or tumor-specific polynucleotides L552S, L550S, L762P, L984P, L763P and L587 having sequences set forth in SEQ ID NOs: 1, 3, 5, 7, 21 and 26, illustrative polypeptides encoded thereby having amino acid sequences set forth in

SEQ ID NO: 2, 4, 6, 8, 22 and 27 that may be suitably employed in the detection of cancer, more specifically, lung cancer.

Methodologies for the Detection of Cancer

- 5           In general, a cancer cell may be detected in a patient based on the presence of one or more polynucleotides within cells of a biological sample (for example, blood, lymph nodes, bone marrow, sera, sputum, urine and/or tumor biopsies) obtained from the patient. In other words, such polynucleotides may be used as markers to indicate the presence or absence of a cancer such as, *e.g.*, lung cancer.
- 10           As discussed in further detail herein, the present invention achieves these and other related objectives by providing a methodology for the simultaneous detection of more than one polynucleotide, the presence of which is diagnostic of the presence of lung cancer cells in a biological sample. Each of the various cancer detection methodologies disclosed herein have in common a step of hybridizing one or more oligonucleotide primers and/or
- 15 probes, the hybridization of which is demonstrative of the presence of a tumor- and/or tissue-specific polynucleotide. Depending on the precise application contemplated, it may be preferred to employ one or more intron-spanning oligonucleotides that are inoperative against polynucleotide of genomic DNA and, thus, these oligonucleotides are effective in substantially reducing and/or eliminating the detection of genomic DNA in the biological
- 20 sample.

Further disclosed herein are methods for enhancing the sensitivity of these detection methodologies by subjecting the biological samples to be tested to one or more cell capture and/or cell depletion methodologies.

- 25           By certain embodiments of the present invention, the presence of lung cancer cell in a patient may be determined by employing the following steps: (a) contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins as described herein; (b) detecting in the sample a level of at least one of the polynucleotides (such as, for example, mRNA) that hybridize to the oligonucleotides; and (c) comparing the level of

polynucleotides that hybridize to the oligonucleotides with a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer in the patient.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a lung tumor protein that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence recited in SEQ ID NO: 1, 3, 5 or 7. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

The present invention also provides amplification-based methods for detecting the presence of lung cancer cells in a patient. Exemplary methods comprise the steps of (a) obtaining a biological sample from the patient; (b) contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof (c) amplifying the polynucleotides; and (d) detecting the amplified polynucleotides; wherein the presence of one or more of the amplified polynucleotides indicates the presence of lung cancer cells in the patient.

Methods according to the present invention are suitable for identifying polynucleotides obtained from a wide variety of biological sample such as, for example, blood, serum, lymph node, bone marrow, sputum, urine and tumor biopsy sample, among others.

Certain exemplary embodiments of the present invention provide methods wherein the polynucleotides to be detected are selected from the group consisting of L762, L984, L550, L552, L763 and L587. Alternatively and/or additionally, polynucleotides to be detected may be selected from the group consisting of those depicted in SEQ ID NOs: 1, 3,  
5 5, 7, 21 and 26.

Suitable exemplary oligonucleotide probes and/or primers that may be used according to the methods of the present invention are disclosed herein. In certain preferred embodiments that eliminate the background detection of genomic DNA, the oligonucleotides may be intron spanning oligonucleotides.

10 Depending on the precise application contemplated, the artisan may prefer to detect the tissue- and/or tumor-specific polynucleotides by detecting a radiolabel and detecting a fluorophore. More specifically, the oligonucleotide probe and/or primer may comprises a detectable moiety such as, for example, a radiolabel and/or a fluorophore.

Alternatively or additionally, methods of the present invention may also comprise a  
15 step of fractionation prior to detection of the tissue- and/or tumor-specific polynucleotides such as, for example, by gel electrophoresis.

In other embodiments, methods described herein may be used as to monitor the progression of cancer. By these embodiments, assays as provided for the diagnosis of lung cancer may be performed over time, and the change in the level of reactive polypeptide(s)  
20 or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

25 Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple lung tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based  
5 on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

#### Cell Enrichment

In other aspects of the present invention, cell capture technologies may be used  
10 prior to polynucleotide detection to improve the sensitivity of the various detection methodologies disclosed herein.

Exemplary cell enrichment methodologies employ immunomagnetic beads that are coated with specific monoclonal antibodies to surface cell markers, or tetrameric antibody complexes, may be used to first enrich or positively select cancer cells in a sample.  
15 Various commercially available kits may be used, including Dynabeads® Epithelial Enrich (DynaL Biotech, Oslo, Norway), StemSep™ (StemCell Technologies, Inc., Vancouver, BC), and RosetteSep (StemCell Technologies). The skilled artisan will recognize that other readily available methodologies and kits may also be suitably employed to enrich or positively select desired cell populations.

20 Dynabeads® Epithelial Enrich contains magnetic beads coated with mAbs specific for two glycoprotein membrane antigens expressed on normal and neoplastic epithelial tissues. The coated beads may be added to a sample and the sample then applied to a magnet, thereby capturing the cells bound to the beads. The unwanted cells are washed away and the magnetically isolated cells eluted from the beads and used in further analyses.

25 RosetteSep can be used to enrich cells directly from a blood sample and consists of a cocktail of tetrameric antibodies that target a variety of unwanted cells and crosslinks them to glycophorin A on red blood cells (RBC) present in the sample, forming rosettes. When centrifuged over Ficoll, targeted cells pellet along with the free RBC.

The combination of antibodies in the depletion cocktail determines which cells will be removed and consequently which cells will be recovered. Antibodies that are available include, but are not limited to: CD2, CD3, CD4, CD5, CD8, CD10, CD11b, CD14, CD15, CD16, CD19, CD20, CD24, CD25, CD29, CD33, CD34, CD36, CD38, CD41, CD45, 5 CD45RA, CD45RO, CD56, CD66B, CD66e, HLA-DR, IgE, and TCR $\alpha\beta$ . Additionally, it is contemplated in the present invention that mAbs specific for lung tumor antigens, can be developed and used in a similar manner. For example, mAbs that bind to tumor-specific cell surface antigens may be conjugated to magnetic beads, or formulated in a tetrameric antibody complex, and used to enrich or positively select metastatic lung tumor cells from 10 a sample. Such a system can be used to evaluate blood samples from different forms of lung cancers, in particular adneo and squamous forms of NSCLC and small cell carcinomas for the presence of circulating tumor cells using the inventive multiplex PCR assay as described herein.

Once a sample is enriched or positively selected, cells may be further analyzed. For 15 example, the cells may be lysed and RNA isolated. RNA may then be subjected to RT-PCR analysis using lung tumor-specific multiplex primers in a Real-time PCR assay as described herein.

In another aspect of the present invention, cell capture technologies may be used in conjunction with Real-Time PCR to provide a more sensitive tool for detection of 20 metastatic cells expressing lung tumor antigens.

Yet another method that may be employed is an anti-ganglioside G<sub>M1</sub>/G<sub>M1</sub> cell capture antibody system. Gangliosides are cell membrane bound glycosphingolipids, several species of which have been shown to be over-expressed on the cell surface of most cancers of neuroectodermal and epithelial origin, in particular lung cancer. Cell surface 25 expression of G<sub>M2</sub> is seen in several types of lung cancer, particularly in SCLC which make it an attractive target for a monoclonal antibody based lung cancer immunotherapy and also for use as a capture method in conjunction with G<sub>M1</sub>.

#### Probes and Primers

As noted above and as described in further detail herein, certain methods, compositions and kits according to the present invention utilize two or more oligonucleotide primer pairs for the detection of lung cancer. The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use  
5 in detecting the presence of complementary sequences in a biological sample.

Alternatively, in other embodiments, the probes and/or primers of the present invention may be employed for detection via nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is  
10 complementary to, a 15 nucleotide long contiguous sequence of a polynucleotide to be detected will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

15 Oligonucleotide primers having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide to be detected, are particularly contemplated as primers for use in amplification reactions such as, *e.g.*, the polymerase chain reaction (PCR<sup>TM</sup>). This would allow a polynucleotide to be  
20 analyzed, both in diverse biological samples such as, for example, blood, lymph nodes and bone marrow.

The use of a primer of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally  
25 preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design primers having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Primers may be selected from any portion of the polynucleotide to be detected. All that is required is to review the sequence, such as those exemplary polynucleotides set forth herein or to any continuous portion of the sequence, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a primer. The choice of primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence. The exemplary primers disclosed herein may optionally be used for their ability to selectively form duplex molecules with complementary stretches of the entire polynucleotide of interest such as those set forth SEQ ID NOs: 1, 3, 5, 7, 21 and 26.

10 The present invention further provides the nucleotide sequence of various exemplary oligonucleotide primers and probes, that may be used, as described in further detail herein, according to the methods of the present invention for the detection of cancer.

Oligonucleotide primers according to the present invention may be readily prepared routinely by methods commonly available to the skilled artisan including, for example, 15 directly synthesizing the primers by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, 20 one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

## 25 Polynucleotide Amplification Techniques

Each of the specific embodiments outlined herein for the detection of lung cancer has in common the detection of a tissue- and/or tumor-specific polynucleotide via the hybridization of one or more oligonucleotide primers and/or probes. Depending on such factors as the relative number of cancer cells present in the biological sample and/or the



level of polynucleotide expression within each lung cancer cell, it may be preferred to perform an amplification step prior to performing the steps of detection. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a lung tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the lung tumor protein. The amplified cDNA may optionally be subjected to a fractionation step such as, for example, gel electrophoresis.

A number of template dependent processes are available to amplify the target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR<sup>TM</sup>) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159. Briefly, in PCR<sup>TM</sup>, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR<sup>TM</sup> amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

One preferred methodology for polynucleotide amplification employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as blood, serum, lymph node, bone marrow, sputum, urine and tumor biopsy samples, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed

on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

Any of a variety of commercially available kits may be used to perform the  
5 amplification step. One such amplification technique is inverse PCR (*see* Triglia *et al.*,  
*Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in  
the known region of the gene. The fragment is then circularized by intramolecular ligation  
and used as a template for PCR with divergent primers derived from the known region.  
Within an alternative approach, sequences adjacent to a partial sequence may be retrieved  
10 by amplification with a primer to a linker sequence and a primer specific to a known  
region. The amplified sequences are typically subjected to a second round of amplification  
with the same linker primer and a second primer specific to the known region. A variation  
on this procedure, which employs two primers that initiate extension in opposite directions  
from the known sequence, is described in WIPO International Patent Application No.: WO  
15 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or  
RACE. This technique involves the use of an internal primer and an external primer,  
which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and  
3' of a known sequence. Additional techniques include capture PCR (Lagerstrom *et al.*,  
*PCR Methods Applic.* 1:111-19, 1991) and walking PCR (Parker *et al.*, *Nucl. Acids. Res.*  
20 19:3055-60, 1991). Other methods employing amplification may also be employed to  
obtain a full length cDNA sequence.

Another method for amplification is the ligase chain reaction (referred to as LCR),  
disclosed in Eur. Pat. Appl. Publ. No. 320,308. In LCR, two complementary probe pairs  
are prepared, and in the presence of the target sequence, each pair will bind to opposite  
25 complementary strands of the target such that they abut. In the presence of a ligase, the  
two probe pairs will link to form a single unit. By temperature cycling, as in PCR™, bound  
ligated units dissociate from the target and then serve as "target sequences" for ligation of  
excess probe pairs. U.S. Patent No. 4,883,750, describes an alternative method of  
amplification similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the  
5 replicative sequence that can then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[ $\alpha$ -thio]triphosphates in one strand of a restriction site (Walker *et al.*, 1992), may also be useful in the amplification of nucleic acids in the present invention.

10 Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.* nick translation. A similar method, called Repair Chain Reaction (RCR) is another method of amplification which may be useful in the present invention and is involves annealing several probes throughout a region targeted for  
15 amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA.

Sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having a 3' and 5' sequences of non-target DNA and an internal or "middle"  
20 sequence of the target protein specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNaseH, and the products of the probe are identified as distinctive products by generating a signal that is released after digestion. The original template is annealed to another cycling probe and the reaction is repeated. Thus, CPR involves amplifying a signal generated by hybridization of a probe to  
25 a target gene specific expressed nucleic acid.

Still other amplification methods described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template and enzyme dependent synthesis. The primers may be modified by

labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes is added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe  
5 signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (Kwoh *et al.*, 1989; PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. In NASBA, the nucleic acids can be prepared for amplification by standard  
10 phenol/chloroform extraction, heat denaturation of a sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has sequences specific to the target sequence. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat-denatured again. In either  
15 case the single stranded DNA is made fully double stranded by addition of second target-specific primer, followed by polymerization. The double stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNAs are reverse transcribed into DNA, and transcribed once again with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target-  
20 specific sequences.

Eur. Pat. Appl. Publ. No. 329,822, disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a first template for a first primer oligonucleotide, which is  
25 elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in a duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to its

template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting as a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence  
5 can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence can be chosen to be in the form of either DNA or RNA.

10 PCT Intl. Pat. Appl. Publ. No. WO 89/06700, disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic; *i.e.* new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" (Frohman, 1990), and  
15 "one-sided PCR" (Ohara, 1989) which are well-known to those of skill in the art.

#### Compositions and Kits for the Detection of Cancer

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers  
20 and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a lung tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection  
25 reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

The present invention also provides kits that are suitable for performing the detection methods of the present invention. Exemplary kits comprise oligonucleotide

primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Within certain embodiments, kits according to the present invention may also comprise a nucleic acid polymerase and suitable buffer. Exemplary oligonucleotide primers suitable for kits of the present invention are disclosed herein. Exemplary polynucleotides suitable for kits  
5 of the present invention are disclosed herein.

Alternatively, a kit may be designed to detect the level of mRNA encoding a lung tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a lung tumor protein. Such an oligonucleotide may be used, for example, within  
10 a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a lung tumor protein.

In other related aspects, the present invention further provides compositions useful in the methods disclosed herein. Exemplary compositions comprise two or more  
15 oligonucleotide primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Exemplary oligonucleotide primers suitable for compositions of the present invention are disclosed herein. Exemplary polynucleotides suitable for compositions of the present invention are disclosed herein.

The following Example is offered by way of illustration and not by way of  
20 limitation.

## EXAMPLES

## EXAMPLE 1

## MULTIPLEX DETECTION OF LUNG TUMORS

A Multiplex Real-time PCR assay was established in order to simultaneously detect the expression of four lung cancer-specific genes: L762 (SEQ ID NO:1), L984 (SEQ ID NO:3), L550 (SEQ ID NO:5) and L552 (SEQ ID NO:7). In contrast to detection approaches relying on expression analysis of single lung cancer-specific genes, this Multiplex assay was able to detect all lung tumor samples tested and analyze their combined mRNA expression profile in adenocarcinoma, squamous, small cell and large cell lung tumors. L552S and L550S complement each other in detecting predominantly adenocarcinomas, L762S detects squamous cell carcinomas and L984P detects small cell carcinomas (see Table 1).

The primers and probes were designed to be intron spanning (exon specific) to eliminate any reactivity with genomic DNA making them suitable for use in blood samples without having to DNase treat mRNA samples. They were also designed to produce amplicons of different sizes to allow gel differentiation of end products if necessary.

The assay was carried out as follows: L552S (SEQ ID NO: 7), L550 (SEQ ID NO: 5), L762 (SEQ ID NO: 1), L984 (SEQ ID NO: 3) and specific primers, and specific Taqman probes, were used to analyze their combined mRNA expression profile in lung tumors. The primers and probes are shown below:

**L552S:** Forward Primer (SEQ ID NO:9): 5' GACGGCATGAGCGACACACA. Reverse Primer (SEQ ID NO:10): 5' CCATGTCGCGCACTGGGATC. Probe (SEQ ID NO:11) (FAM-5' – 3'-TAMRA): CTGAAAGTCGGGATCCTACACCTGGGCA.

**L550P:** Forward Primer (SEQ ID NO:12): 5' GGCCACCGTCTGGATTCTTC. Reverse Primer (SEQ ID NO:13): 5' GAAGAATCCAGACGGTGGCC. Probe (SEQ ID NO:14) (FAM-5' – 3'-TAMRA): CCGCCCCAAG ATCAAATCCA CAAACC.

**L762S:** Forward Primer (SEQ ID NO:15): 5' ATGGCAGAGGCTGACAGACTC. Reverse Primer (SEQ ID NO:16): 5' TTCAACCACCTCAAATCCTTTCTTA. Probe (SEQ ID NO:17) (FAM-5' – 3'-TAMRA) TCGACAGCAAAGGAGAGATCAGAGCCC.

- 5 **L984P:** Forward Primer (SEQ ID NO:18): 5' TTACGACCCGCTCAGCCC. Reverse Primer (SEQ ID NO:19): 5' CTCCCAACGCCACTGACAA. Probe (SEQ ID NO:20) (FAM-5' – 3'-TAMRA): CCAGGCCGAGCCCCTCAGAACC.

The assay conditions were:

*Taqman protocol (7700 Perkin Elmer):*

- 10 In 25 µl final volume: 1x Buffer A, 5mM MgCl, 0.2 mM dCTP, 0.2 mM dATP, 0.4 mM dUTP, 0.2 mM dGTP, 0.01 U/µl AmpErase UNG, 0.0375 U/µl TaqGold, 8% (v/v) Glycerol, 0.05% (v/v) (Sigma), Gelatin, 0.05% (v/v) (Sigma), Tween20 0.1% v/v (Sigma), 300mM of each forward and reverse primer for L762P, 50mM of each forward and reverse primer from (L552S, L984P, L550S, L984P) 2 pmol of each gene specific Taqman probe  
15 (L552S, L550S, L984P) and template cDNA. The PCR reaction was carried out at one cycle at 95°C for 10 minutes, followed by 50 cycles at 95°C for 15 seconds, 60°C for 1 minute, and 68°C for 1 minute (ABI Prism 7900HO Sequence Detection System, Foster City, CA).

- Since each primer set in the multiplex assay results in a band of unique length,  
20 expression signals of the four genes of interest was measured individually by agarose gel analysis. The combined expression signal of all four genes can also be measured in real-time on an ABI 7700 Prism sequence detection system (Applied Biosystems, Foster City, CA). Although specific primers have been described herein, different primer sequences, different primer or probe labeling and different detection systems could be used to perform  
25 this multiplex assay. For example, a second fluorogenic reporter dye could be incorporated for parallel detection of a reference gene by real-time PCR. Or, for example a SYBR Green detection system could be used instead of the Taqman probe approach. Table 2 shows the reactivity of the multiplex PCR with different lung tumor types and normal lung tissue.



**TABLE 2 Expression of Lung Cancer Multiplex Genes (L762P, L552S, L550S, L984P) in Lung Tumor and Normal Lung**

<b>Lung Tumor Type</b>	<b>Positive Samples/Samples Tested</b>
Adenocarcinoma	21/24
Squamous	17/18
Large Cell	5*/5
Small Cell	5/6
Normal Lung Tissue	0/12
<b>Total Tumors</b>	<b>48/53</b>
<b>% Positive Tumors</b>	<b>90.57%</b>

5      Cut-off Value = Mean normal lung +3 SD =0.901

\* One sample at cut-off

## **EXAMPLE 2**

### **MULTIPLEX DETECTION OF LUNG TUMORS**

10

Six additional Multiplex Real-time PCR assays were established in order to simultaneously detect the expression of various combinations of recognized lung antigens: L762 (SEQ ID NO:1), L984 (SEQ ID NO:3), L550 (SEQ ID NO:5), L552 (SEQ ID NO:7), L763 (SEQ ID NO: 21) and L587 (SEQ ID NO:26). The six groups consisted of:

15

Group 1: L762, L552, L550 and L984

Group 2: L763, L552, L550 and L984

Group 3: L763, L552, L587 and L984

Group 4: L763, L550, L587 and L984

Group 5: L763, L550 and L587

20

Group 6: L762, L984, L550 and L587

The assays were carried out described above in Example 1 to analyze the combined mRNA expression profile in lung tumors. The primers and probes for L552S, L550P,

L762S, L984P are as described in Example 1. primers and probes for L763 and L587 are described below:

**L763S:** Forward Primer (SEQ ID NO:23): 5' ATTCCAGGCGACATCCTCACT.  
Reverse Primer (SEQ ID NO:24): 5' GTTTATCCCTGAGTCCTGTTTCCA. Probe (SEQ  
5 ID NO:25) (FAM-5' – 3'-TAMRA): TGTGCACCATTGGCTTCTAGGCACTCC.

**L587:** Forward Primer (SEQ ID NO:28): 5' CCCAGAGCTGTGTTAAGGGATC.  
Reverse Primer (SEQ ID NO:29): 5' GTTAAGCGGGATTTCATGTACGA. Probe (SEQ  
10 ID NO:30) (FAM-5' – 3'-TAMRA): AGAACCTGAACCCGTAAAGAAGCCTCCC.

The lung antigens that make up the six multiplex assays are able to detect all lung tumor samples tested and were analyzed for their combined mRNA expression profile in adenocarcinoma, squamous, small cell and large cell lung tumors. The results of these assays is presented in Table 3.

15 **TABLE 3 Expression of Lung Cancer Multiplex Genes in Lung Tumor and Normal Lung**

Lung Tumor Type	Positive Samples/Samples Tested					
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Adenocarcinoma	21/24	21/24	20/24	22/24	22/24	22/24
Squamous	17/18	17/18	18/18	18/18	18/18	18/18
Large Cell	5/5	3/5	4/5	3/5	3/5	4/5
Small Cell	1/2	1/2	1/2	2/2	1/2	2/2
Other	2/2	2/2	2/2	2/2	2/2	2/2
Normal Lung Tissue	0/12	0/12	0/12	0/13	0/13	0/13

<b>Total Tumors</b>	<b>46/51</b>	<b>44/51</b>	<b>45/51</b>	<b>47/51</b>	<b>46/51</b>	<b>48/51</b>
<b>% Positive Tumors</b>	90.20%	86.27%	88.24%	92.16%	90.20%	94.12%
	CO= 0.9	CO=4.7	CO=1.08	CO=1.88	CO=2.2	CO=5.5

Cut-off Value (CO) = Mean normal lung +3 SD

Multiplex assays using groups 1, 4 and 6 were next used to detect circulating tumor cells in peripheral blood samples from 17 lung cancer patients undergoing various types of treatments. In addition, a single gene assay using lung antigen L523 (SEQ ID NO:31) was carried out in parallel using the primers as described in SEQ ID NOs:33 and 34. Six normal donors were included as controls. The assays were carried out as described above in Example 1. The cut off value for detection in the assay being the mean of the normal lung samples + 3 standard deviations.

Group 1 antigens were detected in 5/17 samples tested. Group 4 antigens were detected in 4/17 samples and Group 6 antigens were detected in 8/17 samples. L523 was detected as a single gene in 7/17 samples tested. The combination of antigens in Group 6 was the most sensitive for lung tumor detection in tissue and blood of the groups tested.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

We Claim:

1. A method for detecting the presence of a cancer cell in a patient, said method comprising the steps of:

(a) obtaining a biological sample from said patient;  
(b) contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof;

(c) amplifying said polynucleotides; and

(d) detecting said amplified polynucleotides;

wherein the presence of one or more of said amplified polynucleotides indicates the presence of lung cancer cells in said patient.

2. A method for determining the presence of lung cancer cells in a patient, said method comprising the steps of:

(a) obtaining a biological sample from said patient;

(b) contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins;

(c) detecting in said biological sample an amount of a polynucleotide that hybridizes to at least one of said oligonucleotides; and

(d) comparing the amount of the polynucleotides that hybridizes to said oligonucleotides to a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer cells in the patient.

3. A method for monitoring the progression of lung cancer in a patient, said method comprising the steps of:

- (a) obtaining a first biological sample from said patient;
- (b) contacting said first biological sample with one or more oligonucleotides that hybridize to one or more polynucleotides that encode lung tumor proteins;
- (c) detecting in said first biological sample an amount of at least one of said polynucleotides that hybridize to said oligonucleotides;
- (d) repeating steps (b) and (c) using a second biological sample obtained from said patient at a subsequent point in time; and
- (e) comparing the amount of polynucleotides detected in step (d) with the amount detected in step (c) and therefrom monitoring the progression of lung cancer in said patient.

**METHODS, COMPOSITIONS AND KITS  
FOR THE DETECTION AND MONITORING OF LUNG CANCER**

**ABSTRACT OF THE DISCLOSURE**

Compositions and methods for the diagnosis of lung cancer are disclosed. Such methods are useful to detect early tumors or provide adequate stage/grade information or tumor specificity. Compositions may comprise one or more lung tumor proteins, immunogenic portions thereof, or polynucleotides that encode such portions. Such compositions may be used, for example, to improve lung cancer diagnosis and prognosis and potentially differentiate between NSCLC and SCLC.

## SEQUENCE LISTING

<110> Zehentner-Wilkinson, Barbara K.  
 Hayes, Dawn  
 Houghton, Raymond L.

<120> METHODS, COMPOSITIONS AND KITS FOR THE DETECTION  
 AND MONITORING OF LUNG CANCER

<130> 609P2

<140> US

<141> 2003-09-15

<160> 34

<170> Corixa Invention Disclosure Database

<210> 1

<211> 3951

<212> DNA

<213> Homo sapiens

<400> 1

tctgcatcca	tattgaaaac	ctgacacaat	gtatgcagca	ggctcagtgt	gagtgaactg	60
gaggcttctc	tacaacatga	cccaaaggag	cattgcaggt	cctatttgca	acctgaagtt	120
tgtgactctc	ctgggtgcct	taagttcaga	actcccattc	ctgggagctg	gagtacagct	180
tcaagacaat	gggtataatg	gattgctcat	tgcaattaat	cctcaggtag	ctgagaatca	240
gaacctcatc	tcaaacatta	aggaaatgat	aactgaagct	tcattttacc	tatttaatgc	300
taccaagaga	agagtatttt	tcagaaatat	aaagatttta	atacctgcca	catggaaagc	360
taataataac	agcaaaataa	aacaagaatc	atatgaaaag	gcaaatgtca	tagtgactga	420
ctgggtatggg	gcacatggag	atgatccata	caccctacaa	tacagagggg	gtggaaaaga	480
gggaaaatac	attcattttca	cacctaat	cctactgaat	gataacttaa	cagctgggta	540
cggatcacga	ggccgagtgt	ttgtccatga	atgggcccac	ctccgttggg	gtgtgttcga	600
tgagtataac	aatgacaaac	ctttctacat	aaatgggcaa	aatcaaatta	aagtgacaag	660
gtgttcacat	gacatcacag	gcatttttgt	gtgtgaaaaa	ggtccttgcc	cccaagaaaa	720
ctgtattatt	agtaagcttt	ttaaagaagg	atgcaccttt	atctacaata	gcacccaaaa	780
tgcaactgca	tcaataatgt	tcatgcaaag	tttatcttct	gtgggttgat	tttgtaatgc	840
aagtaccac	aaccaagaag	caccaaacct	acagaaccag	atgtgcagcc	tcagaagtgc	900
atgggatgta	atcacagact	ctgctgactt	tcaccacagc	tttcccatga	acgggactga	960
gcttccacct	cctcccat	tctcgcttgt	agaggctggg	gacaaagtgg	tctgtttagt	1020
gctggatgtg	tccagcaaga	tggcagaggc	tgacagactc	cttcaactac	aacaagccgc	1080
agaattttat	ttgatgcaga	ttgttgaaat	tcataccttc	gtgggcattg	ccagtttcga	1140
cagcaaagga	gagatcagag	cccagctaca	ccaaattaac	agcaatgatg	atcgaaagtt	1200
gctgggtttca	tatctgcca	ccactgtatc	agctaaaaca	gacatcagca	tttgttcagg	1260
gcttaagaaa	ggatttgagg	tggttgaaaa	actgaatgga	aaagcttatg	gctctgtgat	1320
gatatttagtg	accagcggag	atgataagct	tcttggcaat	tgcttaccca	ctgtgctcag	1380
cagtgggttca	acaattcact	ccattgccct	gggttcacat	gcagcccaa	atctggagga	1440
attatcacgt	cttacaggag	gtttaaagtt	ctttgttcca	gatatatcaa	actccaatag	1500
catgattgat	gctttcagta	gaatttcctc	tggaaactgga	gacattttcc	agcaacatat	1560
tcagcttgaa	agtacagggtg	aaaatgtcaa	acctcaccat	caattgaaaa	acacagtgac	1620
tgtggataat	actgtgggca	acgacactat	gtttctagtt	acgtggcagg	ccagtgggtcc	1680
tcctgagatt	atattatttg	atcctgatgg	acgaaaatac	tacacaaata	attttatcac	1740
caatctaact	tttcggacag	ctagtctttg	gattccagga	acagctaagc	ctgggcactg	1800

```

gacttacacc ctgaacaata cccatcattc tctgcaagcc ctgaaagtga cagtgcctc 1860
tcgcgcctcc aactcagctg tgccccccagc cactgtggaa gcctttgtgg aaagagacag 1920
cctccatttt cctcatcctg tgatgattta tgccaatgtg aaacagggat tttatcccat 1980
tcttaatgcc actgtcactg ccacagttga gccagagact ggagatcctg ttacgctgag 2040
actccttgat gatggagcag gtgctgatgt tataaaaaat gatggaattt actcgaggta 2100
ttttttctcc tttgtgcaa atggtagata tagcttgaaa gtgcatgtca atcactctcc 2160
cagcataaagc accccagccc actctattcc agggagtcat gctatgtatg taccagggtta 2220
cacagcaaac ggtaatatcc agatgaatgc tccaaggaaa tcagtaggca gaaatgagga 2280
ggagcgaaag tggggcttta gccgagtcag ctcaggaggc tccttttcag tgctgggagt 2340
tccagctggc ccccccctg atgtgtttcc accatgcaaa attattgacc tggaaagctgt 2400
aaaagtagaa gaggaattga ccctatcttg gacagcacct ggagaagact ttgatcaggg 2460
ccaggctaca agctatgaaa taagaatgag taaaagtcta cagaatatcc aagatgactt 2520
taacaatgct attttagtaa atacatcaaa gcgaaatcct cagcaagctg gcatcaggga 2580
gatatttacc ttctcacc aaatttcac gaatggacct gaacatcagc caaatggaga 2640
aacacatgaa agccacagaa tttatgttgc aatacgagca atggatagga actccttaca 2700
gtctgtctga tctaaccattg cccaggcgcc tctgtttatt cccccaatt ctgatcctgt 2760
acctgccaga gattatctta tattgaaagg agttttaaca gcaatgggtt tgataggaat 2820
catttgcttt attatagttg tgacacatca tactttaagc aggaaaaaga gagcagacaa 2880
gaaagagaat ggaacaaaat tattataaat aaatatccaa agtgtcttcc ttcttagata 2940
taagacccat ggccttcgac tacaaaaaca tactaacaaa gtcaaattaa catcaaaact 3000
gtattaaaaat gcattgagtt tttgtacaat acagataaga tttttacatg gtagatcaac 3060
aaattctttt tgggggtaga tttagaaaacc cttacacttt ggctatgaac aaataataaa 3120
aattattctt taaagtaatg tctttaaagg caaagggaag ggtaaagtcg gaccagtgtc 3180
aaggaaagtt tgttttattg aggtggaaaa atagcccaa gcagagaaaa ggagggtagg 3240
tctgcattat aactgtctgt gtgaagcaat catttagtta ctttgattaa tttttctttt 3300
ctccttatct gtgcagaaca ggttgcttgt ttacaactga agatcatgct atatttcata 3360
tatgaagccc ctaatgcaaa gctctttacc tcttgctatt ttgttatata tattacagat 3420
gaaatctcac tgctaagtct cagagatctt ttttactgt aagaggtaac ctttaacaat 3480
atgggtatta cttttgtctc ttcataccgg ttttatgaca aaggtctatt gaatttattt 3540
gtttgtaagt ttctactccc atcaaagcag ctttttaagt tattgccttg gttattatgg 3600
atgatagtta tagcccttat aatgccttaa ctaaggaaga aaagatgta ttctgagttt 3660
gttttaatac atatatgaac atatagtttt attcaattaa accaaagaag aggtcagcag 3720
ggagatacta acctttggaa atgattagct ggctctgttt tttgggttaa taagagtctt 3780
taatcctttc tccatcaaga gttacttacc aagggcaggg gaagggggat atagaggtcc 3840
caaggaaata aaaatcatct ttcactctta attttactcc ttcctcttat ttttttaaaa 3900
gattatcgaa caataaaatc atttgctttt ttaattaaaa acataaaaaa a 3951

```

<210> 2

<211> 943

<212> PRT

<213> Homo sapiens

<400> 2

```

Met Thr Gln Arg Ser Ile Ala Gly Pro Ile Cys Asn Leu Lys Phe Val
1          5          10          15
Thr Leu Leu Val Ala Leu Ser Ser Glu Leu Pro Phe Leu Gly Ala Gly
20          25          30
Val Gln Leu Gln Asp Asn Gly Tyr Asn Gly Leu Leu Ile Ala Ile Asn
35          40          45
Pro Gln Val Pro Glu Asn Gln Asn Leu Ile Ser Asn Ile Lys Glu Met
50          55          60
Ile Thr Glu Ala Ser Phe Tyr Leu Phe Asn Ala Thr Lys Arg Arg Val
65          70          75          80
Phe Phe Arg Asn Ile Lys Ile Leu Ile Pro Ala Thr Trp Lys Ala Asn

```



85								90					95				
Asn	Asn	Ser	Lys	Ile	Lys	Gln	Glu	Ser	Tyr	Glu	Lys	Ala	Asn	Val	Ile		
			100					105				110					
Val	Thr	Asp	Trp	Tyr	Gly	Ala	His	Gly	Asp	Asp	Pro	Tyr	Thr	Leu	Gln		
			115				120	125				125					
Tyr	Arg	Gly	Cys	Gly	Lys	Glu	Gly	Lys	Tyr	Ile	His	Phe	Thr	Pro	Asn		
			130			135		140				140					
Phe	Leu	Leu	Asn	Asp	Asn	Leu	Thr	Ala	Gly	Tyr	Gly	Ser	Arg	Gly	Arg		
145				150				155			160						
Val	Phe	Val	His	Glu	Trp	Ala	His	Leu	Arg	Trp	Gly	Val	Phe	Asp	Glu		
			165					170			175						
Tyr	Asn	Asn	Asp	Lys	Pro	Phe	Tyr	Ile	Asn	Gly	Gln	Asn	Gln	Ile	Lys		
			180				185	190			190						
Val	Thr	Arg	Cys	Ser	Ser	Asp	Ile	Thr	Gly	Ile	Phe	Val	Cys	Glu	Lys		
			195			200		205			205						
Gly	Pro	Cys	Pro	Gln	Glu	Asn	Cys	Ile	Ile	Ser	Lys	Leu	Phe	Lys	Glu		
			210			215		220			220						
Gly	Cys	Thr	Phe	Ile	Tyr	Asn	Ser	Thr	Gln	Asn	Ala	Thr	Ala	Ser	Ile		
225				230				235			240						
Met	Phe	Met	Gln	Ser	Leu	Ser	Ser	Val	Val	Glu	Phe	Cys	Asn	Ala	Ser		
			245					250			255						
Thr	His	Asn	Gln	Glu	Ala	Pro	Asn	Leu	Gln	Asn	Gln	Met	Cys	Ser	Leu		
			260				265	270			270						
Arg	Ser	Ala	Trp	Asp	Val	Ile	Thr	Asp	Ser	Ala	Asp	Phe	His	His	Ser		
			275			280		285			285						
Phe	Pro	Met	Asn	Gly	Thr	Glu	Leu	Pro	Pro	Pro	Pro	Thr	Phe	Ser	Leu		
			290			295		300			300						
Val	Glu	Ala	Gly	Asp	Lys	Val	Val	Cys	Leu	Val	Leu	Asp	Val	Ser	Ser		
305				310				315			320						
Lys	Met	Ala	Glu	Ala	Asp	Arg	Leu	Leu	Gln	Leu	Gln	Gln	Ala	Ala	Glu		
			325				330	335			335						
Phe	Tyr	Leu	Met	Gln	Ile	Val	Glu	Ile	His	Thr	Phe	Val	Gly	Ile	Ala		
			340				345	350			350						
Ser	Phe	Asp	Ser	Lys	Gly	Glu	Ile	Arg	Ala	Gln	Leu	His	Gln	Ile	Asn		
			355			360		365			365						
Ser	Asn	Asp	Asp	Arg	Lys	Leu	Leu	Val	Ser	Tyr	Leu	Pro	Thr	Thr	Val		
			370			375		380			380						
Ser	Ala	Lys	Thr	Asp	Ile	Ser	Ile	Cys	Ser	Gly	Leu	Lys	Lys	Gly	Phe		
385				390				395			400						
Glu	Val	Val	Glu	Lys	Leu	Asn	Gly	Lys	Ala	Tyr	Gly	Ser	Val	Met	Ile		
			405				410	415			415						
Leu	Val	Thr	Ser	Gly	Asp	Asp	Lys	Leu	Leu	Gly	Asn	Cys	Leu	Pro	Thr		
			420				425	430			430						
Val	Leu	Ser	Ser	Gly	Ser	Thr	Ile	His	Ser	Ile	Ala	Leu	Gly	Ser	Ser		
			435			440		445			445						
Ala	Ala	Pro	Asn	Leu	Glu	Glu	Leu	Ser	Arg								

		515					520					525				
Thr	Trp	Gln	Ala	Ser	Gly	Pro	Pro	Glu	Ile	Ile	Leu	Phe	Asp	Pro	Asp	
	530					535					540					
Gly	Arg	Lys	Tyr	Tyr	Thr	Asn	Asn	Phe	Ile	Thr	Asn	Leu	Thr	Phe	Arg	
545					550					555					560	
Thr	Ala	Ser	Leu	Trp	Ile	Pro	Gly	Thr	Ala	Lys	Pro	Gly	His	Trp	Thr	
				565					570					575		
Tyr	Thr	Leu	Asn	Asn	Thr	His	His	Ser	Leu	Gln	Ala	Leu	Lys	Val	Thr	
			580					585					590			
Val	Thr	Ser	Arg	Ala	Ser	Asn	Ser	Ala	Val	Pro	Pro	Ala	Thr	Val	Glu	
		595					600					605				
Ala	Phe	Val	Glu	Arg	Asp	Ser	Leu	His	Phe	Pro	His	Pro	Val	Met	Ile	
	610					615					620					
Tyr	Ala	Asn	Val	Lys	Gln	Gly	Phe	Tyr	Pro	Ile	Leu	Asn	Ala	Thr	Val	
625					630					635					640	
Thr	Ala	Thr	Val	Glu	Pro	Glu	Thr	Gly	Asp	Pro	Val	Thr	Leu	Arg	Leu	
				645					650					655		
Leu	Asp	Asp	Gly	Ala	Gly	Ala	Asp	Val	Ile	Lys	Asn	Asp	Gly	Ile	Tyr	
			660					665					670			
Ser	Arg	Tyr	Phe	Phe	Ser	Phe	Ala	Ala	Asn	Gly	Arg	Tyr	Ser	Leu	Lys	
		675					680					685				
Val	His	Val	Asn	His	Ser	Pro	Ser	Ile	Ser	Thr	Pro	Ala	His	Ser	Ile	
	690					695					700					
Pro	Gly	Ser	His	Ala	Met	Tyr	Val	Pro	Gly	Tyr	Thr	Ala	Asn	Gly	Asn	
705					710					715					720	
Ile	Gln	Met	Asn	Ala	Pro	Arg	Lys	Ser	Val	Gly	Arg	Asn	Glu	Glu	Glu	
				725					730					735		
Arg	Lys	Trp	Gly	Phe	Ser	Arg	Val	Ser	Ser	Gly	Gly	Ser	Phe	Ser	Val	
			740					745					750			
Leu	Gly	Val	Pro	Ala	Gly	Pro	His	Pro	Asp	Val	Phe	Pro	Pro	Cys	Lys	
		755					760				765					
Ile	Ile	Asp	Leu	Glu	Ala	Val	Lys	Val	Glu	Glu	Glu	Leu	Thr	Leu	Ser	
	770					775					780					
Trp	Thr	Ala	Pro	Gly	Glu	Asp	Phe	Asp	Gln	Gly	Gln	Ala	Thr	Ser	Tyr	
785					790					795					800	
Glu	Ile	Arg	Met	Ser	Lys	Ser	Leu	Gln	Asn	Ile	Gln	Asp	Asp	Phe	Asn	
				805					810					815		
Asn	Ala	Ile	Leu	Val	Asn	Thr	Ser	Lys	Arg	Asn	Pro	Gln	Gln	Ala	Gly	
			820					825					830			
Ile	Arg	Glu	Ile	Phe	Thr	Phe	Ser	Pro	Gln	Ile	Ser	Thr	Asn	Gly	Pro	
		835					840					845				
Glu	His	Gln	Pro	Asn	Gly	Glu	Thr	His	Glu	Ser	His	Arg	Ile	Tyr	Val	
	850					855					860					
Ala	Ile	Arg	Ala	Met	Asp	Arg	Asn	Ser	Leu	Gln	Ser	Ala	Val	Ser	Asn	
865					870					875						

<210> 3  
 <211> 785  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 tctgattccg cgactccttg gccgcccgtg cgcattggaaa gctctgccaa gatggagagc 60  
 ggccggcgccg gccagcagcc ccagccgcag ccccagcagc ccttcctgcc gcccgagcc 120  
 tgtttctttg ccacggccgc agccgcggcg gccgcagccg ccgcagcggc agcgagagc 180  
 gcgcagcagc agcagcagca gcagcagcag caggcgccgc agctgagacc ggccggccgac 240  
 ggccagccct cagggggcg tacaagtca gcgcccaagc aagtcaagcg acagcgctcg 300  
 tcttcgcccg aactgatgag ctgcaaagcg cggctcaact tcagcggctt tggctacagc 360  
 ctgccgcagc agcagccggc cgcctgtggc cgcgcgaac agcgcgagcg caaccgcgtc 420  
 aagttgggtc acctgggctt tgccaccctt cgggagcagc tccccaacgg cgcggccaac 480  
 aagaagatga gtaaggtgga gacactgcgc tcggcggctg agtacatccg cgcgctgcag 540  
 cagctgctgg acgagcatga cgcggtgagc gccgccttcc aggcaggcgt cctgtcgcgc 600  
 accatctccc ccaactactc caacgacttg aactccatgg ccggctcgcc ggtctcatcc 660  
 tactcgtcgg acgagggctc ttacgaccgc ctacgccccg aggagcagga gcttctcgac 720  
 ttcaccaact ggttctgagg ggctcggcct ggtcaggccc tgggtcgaat ggactttgga 780  
 agcag 785

<210> 4  
 <211> 236  
 <212> PRT  
 <213> Homo sapiens

<400> 4  
 Met Glu Ser Ser Ala Lys Met Glu Ser Gly Gly Ala Gly Gln Gln Pro  
 1 5 10 15  
 Gln Pro Gln Pro Gln Gln Pro Phe Leu Pro Pro Ala Ala Cys Phe Phe  
 20 25 30  
 Ala Thr Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gln  
 35 40 45  
 Ser Ala Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ala Pro  
 50 55 60  
 Gln Leu Arg Pro Ala Ala Asp Gly Gln Pro Ser Gly Gly Gly His Lys  
 65 70 75 80  
 Ser Ala Pro Lys Gln Val Lys Arg Gln Arg Ser Ser Ser Pro Glu Leu  
 85 90 95  
 Met Arg Cys Lys Arg Arg Leu Asn Phe Ser Gly Phe Gly Tyr Ser Leu  
 100 105 110  
 Pro Gln Gln Gln Pro Ala Ala Val Ala Arg Arg Asn Glu Arg Glu Arg  
 115 120 125  
 Asn Arg Val Lys Leu Val Asn Leu Gly Phe Ala Thr Leu Arg Glu His  
 130 135 140  
 Val Pro Asn Gly Ala Ala Asn Lys Lys Met Ser Lys Val Glu Thr Leu  
 145 150 155 160  
 Arg Ser Ala Val Glu Tyr Ile Arg Ala Leu Gln Gln Leu Leu Asp Glu  
 165 170 175  
 His Asp Ala Val Ser Ala Ala Phe Gln Ala Gly Val Leu Ser Pro Thr  
 180 185 190  
 Ile Ser Pro Asn Tyr Ser Asn Asp Leu Asn Ser Met Ala Gly Ser Pro  
 195 200 205  
 Val Ser Ser Tyr Ser Ser Asp Glu Gly Ser Tyr Asp Pro Leu Ser Pro

210 215 220  
 Glu Glu Gln Glu Leu Leu Asp Phe Thr Asn Trp Phe  
 225 230 235

<210> 5  
 <211> 1633  
 <212> DNA  
 <213> Homo sapiens

<400> 5  
 cgtggaggca gctagcgcga ggctgggggag cgctgagccg cgcgtcgtgc cctgcgctgc 60  
 ccagactagc gaacaataca gtcgggatgg ctaaagggtga cccaagaaa ccaaagggca 120  
 agacgtccgc ttatgccttc tttgtgcaga catgcagaga agaacataag aagaaaaacc 180  
 cagaggtccc tgtcaatttt gcggaatttt ccaagaagtg ctctgagagg tggaagacgg 240  
 tgtccgggaa agagaaatcc aaatttgatg aaatggcaaa ggcagataaa gtgcgctatg 300  
 atcgggaaat gaaggattat ggaccagcta agggaggcaa gaagaagaag gatcctaata 360  
 ctcccaaaaag gccaccgtct ggattcttcc tgttctgttc agaattccgc cccaagatca 420  
 aatccacaaa ccccggtcatc tctattggag acgtggcaaa aaagctgggt gagatgtgga 480  
 ataatttaaa tgacagtga aagcagcctt acatcactaa ggcggcaaag ctgaaggaga 540  
 agtatgagaa ggatgttgct gactataagt cgaaaggaaa gtttgatggt gcaaaggggtc 600  
 ctgctaaagt tgcccggaaa aaggtggaag aggaagatga agaacaggag gaggaagaag 660  
 aggaggagga ggaggaggag gatgaataaa gaaactgttt atctgtctcc ttgtgaatac 720  
 ttagagtagg ggagcgccgt aattgacaca tctcttattt gagaagtgtc tgttgccctc 780  
 attaggttta attacaaaat ttgatcacga tcatattgta gtctctcaaa gtgctctaga 840  
 aattgtcagt ggtttacatg aagtggccat ggggtgtctgg agcaccctga aactgtatca 900  
 aagttgtaca tatttccaaa catttttaaa atgaaaaggc actctcgtgt tctcctcact 960  
 ctgtgcactt tgctgttggt gtgacaaggc atttaaagat gtttctggca ttttcttttt 1020  
 atttgtaagg tgggtggaac tatggttatt ggctagaaat cctgagtttt caactgtata 1080  
 tatctatagt ttgtaaaaag aacaaaacaa ccgagacaaa cccttgatgc tccttgctcg 1140  
 gcgttgaggc tgtggggaag atgccttttg ggagaggctg tagctcaggc cgtgcactgt 1200  
 gaggtggac ctgttgactc tgcagggggc atccatttag cttcaggttg tcttgtttct 1260  
 gtatatagt acatagcatt ctgctgccat cttagctgtg gacaaagggg ggtcagctgg 1320  
 catgagaata ttttttttta agtgcggtag tttttaaaact gtttgttttt aaacaaacta 1380  
 tagaactctt cattgtcagc aaagcaaaga gtcactgcat caatgaaagt tcaagaacct 1440  
 cctgtactta aacacgattc gcaacgttct gttatttttt ttgtatgttt agaatgctga 1500  
 aatgtttttg aagttaaata aacagtatta catttttaga actcttctct actataacag 1560  
 tcaatttctg actcacagca gtgaacaaac cccactccg ttgtatttgg agactggcct 1620  
 ccctataaat gtg 1633

<210> 6  
 <211> 200  
 <212> PRT  
 <213> Homo sapiens

<400> 6  
 Met Ala Lys Gly Asp Pro Lys Lys Pro Lys Gly Lys Met Ser Ala Tyr  
 1 5 10 15  
 Ala Phe Phe Val Gln Thr Cys Arg Glu His Lys Lys Lys Asn Pro  
 20 25 30  
 Glu Val Pro Val Asn Phe Ala Glu Phe Ser Lys Lys Cys Ser Glu Arg  
 35 40 45  
 Trp Lys Thr Met Ser Gly Lys Glu Lys Ser Lys Phe Asp Glu Met Ala  
 50 55 60  
 Lys Ala Asp Lys Val Arg Tyr Asp Arg Glu Met Lys Asp Tyr Gly Pro



65					70					75				80	
Glu	Ser	Pro	Lys	Lys	Lys	Asn	Gln	Gln	Leu	Lys	Val	Gly	Ile	Leu	His
				85					90					95	
Leu	Gly	Ser	Arg	Gln	Lys	Lys	Ile	Arg	Ile	Gln	Leu	Arg	Ser	Gln	Cys
			100					105					110		
Ala	Thr	Trp	Lys	Val	Ile	Cys	Lys	Ser	Cys	Ile	Ser	Gln	Thr	Pro	Gly
		115					120					125			
Ile	Asn	Leu	Asp	Leu	Gly	Ser	Gly	Val	Lys	Val	Lys	Ile	Ile	Pro	Lys
	130					135					140				
Glu	Glu	His	Cys	Lys	Met	Pro	Glu	Ala	Gly	Glu	Glu	Gln	Pro	Gln	Val
145					150					155					160

<210> 9  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 gacggcatga gcgacacaca

20

<210> 10  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 10  
 ccatgtcgcg cactgggatc

20

<210> 11  
 <211> 28  
 <212> DNA  
 <213> Homo sapiens

<400> 11  
 ctgaaagtcg ggatcctaca cctgggca

28

<210> 12  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 12  
 ggccaccgtc tggattcttc

20

<210> 13  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 13  
 gaagaatcca gacggtggcc

20

<210> 14

<211> 26  
 <212> DNA  
 <213> Homo sapiens

<400> 14  
 ccgcccgaag atcaaattcca caaacc

26

<210> 15  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 15  
 atggcagagg ctgacagact c

21

<210> 16  
 <211> 25  
 <212> DNA  
 <213> Homo sapiens

<400> 16  
 ttcaaccacc tcaaattcctt tctta

25

<210> 17  
 <211> 27  
 <212> DNA  
 <213> Homo sapiens

<400> 17  
 tcgacagcaa aggagagatc agagccc

27

<210> 18  
 <211> 18  
 <212> DNA  
 <213> Homo sapiens

<400> 18  
 ttacgacccg ctcagccc

18

<210> 19  
 <211> 19  
 <212> DNA  
 <213> Homo sapiens

<400> 19  
 ctcccaacgc cactgacaa

19

<210> 20  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens

<400> 20  
 ccaggccgag cccctcagaa cc

22

<210> 21  
 <211> 1800  
 <212> DNA  
 <213> Homo sapiens

<400> 21  
 gcgcctcatt gccactgcag tgactaaagc tgggaagacg ctggtcagtt cacctgcccc 60  
 actgggttgtt ttttaaaca attctgatac aggcgacatc ctactgacc gagcaaagat 120  
 tgacattcgt atcatcactg tgcaccattg gcttctaggc actccagtgg ggtaggagaa 180  
 ggaggtctga aaccctcgca gagggatctt gccctcattc tttgggtctg aaacactggc 240  
 agtcgttgga aacaggactc agggataaac cagcgcaatg gattggggga cgctgcacac 300  
 tttcatcggg ggtgtcaaca aacactccac cagcatcggg aaggtgtgga tcacagtcac 360  
 ctttattttc cgagtcacga tcctagtggg ggctgcccag gaagtgtggg gtgacgagca 420  
 agaggacttc gtctgcaaca cactgcaacc gggatgcaaa aatgtgtgct atgaccactt 480  
 tttcccggtg tcccacatcc ggctgtgggc cctccagctg atcttcgtct ccaccccagc 540  
 gctgctgggt gccatgcatg tggcctacta caggcacgaa accactcgca agttcaggcg 600  
 aggagagaag aggaatgatt tcaaagacat agaggacatt aaaaagcaca aggttcggat 660  
 agaggggtcg ctgtgggtgga cgtacaccag cagcatcttt ttccgaatca tctttgaagc 720  
 agcctttatg tatgtgtttt acttccttta caatgggtac cacctgcccc gggtgttgaa 780  
 atgtgggatt gacccttgcc ccaaccttgt tgactgcttt atttctaggc caacagagaa 840  
 gaccgtgttt accattttta tgatttctgc gtctgtgatt tgcattgctg ttaacgtggc 900  
 agagttgtgc taccgtgtgc tgaaagtgtg ttttaggaga tcaaagagag cacagacgca 960  
 aaaaaatcac cccaatcatg ccctaaagga gagtaagcag aatgaaatga atgagctgat 1020  
 ttcagatagt ggtcaaaatg caatcacagg tttcccaagc taaacatttc aaggtaaaat 1080  
 gtagctgcgt cataaggaga cttctgtctt ctccagaagg caataccaac ctgaaagttc 1140  
 cttctgtagc ctgaagagtt tgtaaatgac tttcataata aatagacact tgagttaact 1200  
 tttttaggga tacttgctcc attcatacac aacgtaatca aatatgtggt ccatctctga 1260  
 aaacaagaga ctgcttgaca aaggagcatt gcagtcactt tgacaggttc cttttaagtg 1320  
 gactctctga caaagtgggt actttctgaa aatttatata actgttggtg ataaggaaca 1380  
 tttatccagg aattgatacg tttattagga aaagatatat ttataggctt ggatgttttt 1440  
 agttccgact ttgaatttat ataaagtatt tttataatga ctggctcttc ttacctggaa 1500  
 aaacatgcga tgtagtattt agaattacac cacaagtatc taaatttcca acttacaaaag 1560  
 ggctctatct tgtaaatatt gttttgcatt gtctgttggc aaatttgtga actgtcatga 1620  
 tacgtttaag gtgggaaggt gttcattgca caatatattt ttactgcttt ctgaatgtag 1680  
 acggaacagt gtggaagcag aaggcttttt taactcatcc gtttggccga tcgttgacga 1740  
 ccactgggag atgtggatgt gggtgcctcc ttttgcctgt ccccggtggt taacctttct 1800

<210> 22  
 <211> 261  
 <212> PRT  
 <213> Homo sapiens

<400> 22  
 Met Asp Trp Gly Thr Leu His Thr Phe Ile Gly Gly Val Asn Lys His  
 1 5 10 15  
 Ser Thr Ser Ile Gly Lys Val Trp Ile Thr Val Ile Phe Ile Phe Arg  
 20 25 30  
 Val Met Ile Leu Val Val Ala Ala Gln Glu Val Trp Gly Asp Glu Gln  
 35 40 45  
 Glu Asp Phe Val Cys Asn Thr Leu Gln Pro Gly Cys Lys Asn Val Cys  
 50 55 60  
 Tyr Asp His Phe Phe Pro Val Ser His Ile Arg Leu Trp Ala Leu Gln  
 65 70 75 80



Leu Ile Phe Val Ser Thr Pro Ala Leu Leu Val Ala Met His Val Ala  
                             85                            90                            95  
 Tyr Tyr Arg His Glu Thr Thr Arg Lys Phe Arg Arg Gly Glu Lys Arg  
                             100                            105                            110  
 Asn Asp Phe Lys Asp Ile Glu Asp Ile Lys Lys His Lys Val Arg Ile  
                             115                            120                            125  
 Glu Gly Ser Leu Trp Trp Thr Tyr Thr Ser Ser Ile Phe Phe Arg Ile  
                             130                            135                            140  
 Ile Phe Glu Ala Ala Phe Met Tyr Val Phe Tyr Phe Leu Tyr Asn Gly  
 145                            150                            155                            160  
 Tyr His Leu Pro Trp Val Leu Lys Cys Gly Ile Asp Pro Cys Pro Asn  
                             165                            170                            175  
 Leu Val Asp Cys Phe Ile Ser Arg Pro Thr Glu Lys Thr Val Phe Thr  
                             180                            185                            190  
 Ile Phe Met Ile Ser Ala Ser Val Ile Cys Met Leu Leu Asn Val Ala  
                             195                            200                            205  
 Glu Leu Cys Tyr Leu Leu Leu Lys Val Cys Phe Arg Arg Ser Lys Arg  
                             210                            215                            220  
 Ala Gln Thr Gln Lys Asn His Pro Asn His Ala Leu Lys Glu Ser Lys  
 225                            230                            235                            240  
 Gln Asn Glu Met Asn Glu Leu Ile Ser Asp Ser Gly Gln Asn Ala Ile  
                             245                            250                            255  
 Thr Gly Phe Pro Ser  
                             260

<210> 23  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 23  
 attccaggcg acatcctcac t

21

<210> 24  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens

<400> 24  
 gtttatccct gagtcctggtt tocca

24

<210> 25  
 <211> 27  
 <212> DNA  
 <213> Homo sapiens

<400> 25  
 tgtgcacat tggcttctag gcactcc

27

<210> 26  
 <211> 2257  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 26

```

atttttgctta cagagtcccg tctcaccatc ctgggcttcc aacggagact gcggtatccg 60
cggctggaga cccagcggcg agtagccttt tgctcccgga cggacttgag aggcttaaag 120
gatggcctcg tcagatctgg aacaattatg ctctcatggt aatgaaaaga ttggcaatat 180
taagaaaacc ttatcattaa gaaactgtgg ccaggaacct accttgaaaa ctgtattaaa 240
taaaatagga gatgagatca ttgtaataaa tgaacttcta aataaattgg aattggaaat 300
tcagtatcaa gaacaaacca acaattcact caagggaactc tgtgaatctc ttgaagaaga 360
ttacaaagac atagaacatc ttaaagaaaa cgttccttcc catttgctc aagtaacagt 420
aaccagagc tgtgttaagg gatcagatct tgatcctgaa gaaccaatca aagttgaaga 480
acctgaacct gtaaagaagc ctcccaaaga gcaaagaagt attaaggaaa tgccatttat 540
aacttgtgat gagttcaatg gtgttccttc gtacatgaaa tcccgcttaa cctataatca 600
aattaatgat gttattaaag aaatcaacaa ggcagtaatt agtaaataa aaatcctaca 660
tcagccaaaa aagtctatga attctgtgac cagaaatctc tatcacagat ttattgatga 720
agaaacgaag gataccaaag gtcgttattt tatagtggaa gctgacataa aggagttcac 780
aactttgaaa gctgacaaga agtttcacgt gttactgaat attttacgac actgccggag 840
gctatcagag gtccgagggg gaggacttac tcgttatggt ataacctgag tcccttgtga 900
acttttgaa ataccaacag ggtatagagt atagaggcta tttctataat tttcttatat 960
ataatttttt taacttttaa tcttttttgt ttcctttttt ttttttttga gacaggatct 1020
tgctttgtca cccaggggct tgctttgtca cgcaggctag agtgcagtgg cgcaaacatg 1080
gctcactgca gctcaacct cccaggctca agtgatcctc ccacctcagc cccctgaatg 1140
gctgggacta caagcgtgcy ccaccatgcc tggctaattt ttgtattttt tggagagatg 1200
gggtttcacc atgttgccca ggctgggtctt gagctcctga gctcaaacaa tccaccctcc 1260
tcagcctccc aaagtgtctg gattacaggc ttgagccacc acacctgacc tattcttggt 1320
tcttataaaa ataaaacttt tttggataaa gcttatttct tgtttttttc tttttctttt 1380
tttttttttt tcgagactcc atctcagaaa aaaagaaaaa aagactgggt acagatgtga 1440
tattggaaga aaaagatcaa gctgatgagg ttaggatacc caggcccttt ggacttaaag 1500
atcactagt tctaaattcc atcgatggca tttcagtcta taggtaaact tcctggaagc 1560
tggatttgga gacagtttat catctgatta ttgggcttcc gtatagggtcc ttagggagca 1620
gcttacctga aatgcattta gtgtacacca gtctgtaaac ttcaacctgt aatgaaagtg 1680
taataaatgt acattgagtt gatgtgataa tgtgatataa taagaaatat atatttgatc 1740
ttcctatcta gttccttggt cagagctcct aaaacccttg taatttccaa agtgatggag 1800
tacatctttt gttctagtat ttgggtcttg accccagttc ctgacacaaa gctcctaaat 1860
tcctttaaat ttcccagtga taggagaatt tttgttcta atgaggtcac tcttgatggg 1920
cacctggata actcaggatg ggggctgctc acaaagacca catcatgatt ggaagtttca 1980
aactttcagt ctcccacctc cagagagggg agaggggctg gagatttggt tcaataatcc 2040
atcaggccta tgtcaacaag acataatccg ttaactatgg agttcaggga gcttcagggt 2100
tggcaaacat tttgatgtgc caggaagggt acgcactcca gctttatgaa gtcagcaagt 2160
cctgtgctca ggatgcttyt ggaccttgcc ccagggtaccc cttcatgtgg ctgttggtca 2220
tctgtatcct ttgtagtagc cttaaaaataa actgtta 2257

```

&lt;210&gt; 27

&lt;211&gt; 255

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 27

```

Met Ala Ser Ser Asp Leu Glu Gln Leu Cys Ser His Val Asn Glu Lys
 1             5             10            15
Ile Gly Asn Ile Lys Lys Thr Leu Ser Leu Arg Asn Cys Gly Gln Glu
          20          25          30
Pro Thr Leu Lys Thr Val Leu Asn Lys Ile Gly Asp Glu Ile Ile Val

```

Ile	Asn	Glu	Leu	Leu	Asn	Lys	Leu	Glu	Leu	Glu	Ile	Gln	Tyr	Gln	Glu
	50					55					60				
Gln	Thr	Asn	Asn	Ser	Leu	Lys	Glu	Leu	Cys	Glu	Ser	Leu	Glu	Glu	Asp
65					70					75					80
Tyr	Lys	Asp	Ile	Glu	His	Leu	Lys	Glu	Asn	Val	Pro	Ser	His	Leu	Pro
				85					90					95	
Gln	Val	Thr	Val	Thr	Gln	Ser	Cys	Val	Lys	Gly	Ser	Asp	Leu	Asp	Pro
			100					105					110		
Glu	Glu	Pro	Ile	Lys	Val	Glu	Glu	Pro	Glu	Pro	Val	Lys	Lys	Pro	Pro
		115				120					125				
Lys	Glu	Gln	Arg	Ser	Ile	Lys	Glu	Met	Pro	Phe	Ile	Thr	Cys	Asp	Glu
	130					135					140				
Phe	Asn	Gly	Val	Pro	Ser	Tyr	Met	Lys	Ser	Arg	Leu	Thr	Tyr	Asn	Gln
145					150					155					160
Ile	Asn	Asp	Val	Ile	Lys	Glu	Ile	Asn	Lys	Ala	Val	Ile	Ser	Lys	Tyr
				165					170					175	
Lys	Ile	Leu	His	Gln	Pro	Lys	Lys	Ser	Met	Asn	Ser	Val	Thr	Arg	Asn
			180					185					190		
Leu	Tyr	His	Arg	Phe	Ile	Asp	Glu	Glu	Thr	Lys	Asp	Thr	Lys	Gly	Arg
		195				200						205			
Tyr	Phe	Ile	Val	Glu	Ala	Asp	Ile	Lys	Glu	Phe	Thr	Thr	Leu	Lys	Ala
	210					215					220				
Asp	Lys	Lys	Phe	His	Val	Leu	Leu	Asn	Ile	Leu	Arg	His	Cys	Arg	Arg
225					230					235					240
Leu	Ser	Glu	Val	Arg	Gly	Gly	Gly	Leu	Thr	Arg	Tyr	Val	Ile	Thr	
				245					250					255	

```
<210> 28
<211> 22
<212> DNA
<213> Homo sapiens
```

<400> 28  
cccagagctg tgттаaggga tc

22

```
<210> 29
<211> 23
<212> DNA
<213> Homo sapiens
```

**<400> 29**  
gttaagcggg atttcattgta cga

23

```
<210> 30
<211> 28
<212> DNA
<213> Homo sapiens
```

**<400> 30**  
**agaacctgaa cccgtaaaga agcctccc**

28

<210> 31  
 <211> 1740  
 <212> DNA  
 <213> Homo sapiens

<400> 31  
 atgaacaaac tgtatatcgg aaacctcagc gagaacgccg cccctcggga cctagaaaagt 60  
 atcttcaagg acgccaagat cccggtgtcg ggacccttcc tgggtgaagac tggctacgcg 120  
 ttcgtggact gcccgacga gagctgggccc ctcaaggcca tcgaggcgct ttcaggtaaa 180  
 atagaactgc acgggaaacc catagaagtt gagcactcgg tcccaaaaag gcaaaggatt 240  
 cggaaacttc agatacgaat tatcccgccct catttacagt gggagggtgct ggatagttta 300  
 ctagtccagt atggagtggt ggagagctgt gaagcaagtga acactgactc ggaaactgca 360  
 gttgtaaatg taacctattc cagtaaggac caagctagac aagcactaga caaactgaat 420  
 ggatttcagt tagagaattt caccttgaaa gtagcctata tccctgatga aacggccgcc 480  
 cagcaaaacc ccttgacgca gcccggaggt cgccgggggc ttgggcagag gggctcctca 540  
 aggcaggggt ctccaggatc cgtatccaag cagaaaccat gtgatttgcc tctgcgcctg 600  
 ctggttccca cccaatttgt tggagccatc ataggaaaag aagggtgccac cattcggaac 660  
 atcaccaaac agacccagtc taaaatcgat gtccaccgta aagaaaatgc gggggctgct 720  
 gagaagtcga ttactatcct ctctactcct gaaggcacct ctgcggcttg taagtctatt 780  
 ctggagatta tgcataagga agctcaagat ataaaattca cagaagagat ccccttgaag 840  
 attttagctc ataataactt tgttggacgt cttattggta aagaaggaag aaatcttaaa 900  
 aaaattgagc aagacacaga cactaaaatc acgatatctc cattgcagga attgacgctg 960  
 tataatccag aacgcactat tacagttaaa ggcaatgttg agacatgtgc caaagctgag 1020  
 gaggagatca tgaagaaaat cagggagctt tatgaaaatg atattgtctt tatgaatctt 1080  
 caagcacatt taattcctgg attaaatctg aacgccttgg gtctgttccc acccacttca 1140  
 gggatgccac ctcccacctc agggcccccct tcagccatga ctctcccta cccgcagttt 1200  
 gagcaatcag aaacggagac tgttcatctg tttatcccag ctctatcagt cggtgccatc 1260  
 atcggaagc agggccagca catcaagcag ctttctcgct ttgctggagc ttcaattaag 1320  
 attgctccag cggaagcacc agatgctaaa gtgaggatgg tgattatcac tggaccacca 1380  
 gaggtcagt tcaaggctca gggagaagatt tatggaaaaa ttaaagaaga aaactttgtt 1440  
 agtcctaaag aagaggtgaa acttgaagct catatcagag tgccatcctt tgctgctggc 1500  
 agagttattg gaaaaggagg caaaacgggtg aatgaacttc agaatttgtc aagtgcagaa 1560  
 gttgttgtcc ctctgaccca gacacctgat gagaatgacc aagtgggtgt caaaataact 1620  
 ggtcacttct atgcttgcca gggtgcccag agaaaaattc aggaattct gactcaggta 1680  
 aagcagcacc aacaacagaa ggctctgcaa agtggaccac ctcagtcaag acggaagtaa 1740

<210> 32  
 <211> 579  
 <212> PRT  
 <213> Homo sapiens

<400> 32  
 Met Asn Lys Leu Tyr Ile Gly Asn Leu Ser Glu Asn Ala Ala Pro Ser  
 1 5 10 15  
 Asp Leu Glu Ser Ile Phe Lys Asp Ala Lys Ile Pro Val Ser Gly Pro  
 20 25 30  
 Phe Leu Val Lys Thr Gly Tyr Ala Phe Val Asp Cys Pro Asp Glu Ser  
 35 40 45  
 Trp Ala Leu Lys Ala Ile Glu Ala Leu Ser Gly Lys Ile Glu Leu His  
 50 55 60  
 Gly Lys Pro Ile Glu Val Glu His Ser Val Pro Lys Arg Gln Arg Ile

65	Arg	Lys	Leu	Gln	Ile	Arg	Asn	Ile	Pro	Pro	75	His	Leu	Gln	Trp	Glu	Val	80
				85							90						95	
Leu	Asp	Ser	Leu	Leu	Val	Gln	Tyr	Gly	Val	Val	Val	Glu	Ser	Cys	Glu	Gln		
			100					105						110				
Val	Asn	Thr	Asp	Ser	Glu	Thr	Ala	Val	Val	Asn	Val	Thr	Tyr	Ser	Ser			
			115				120						125					
Lys	Asp	Gln	Ala	Arg	Gln	Ala	Leu	Asp	Lys	Leu	Asn	Gly	Phe	Gln	Leu			
			130			135					140							
Glu	Asn	Phe	Thr	Leu	Lys	Val	Ala	Tyr	Ile	Pro	Asp	Glu	Thr	Ala	Ala			
				150						155					160			
Gln	Gln	Asn	Pro	Leu	Gln	Gln	Pro	Arg	Gly	Arg	Arg	Gly	Leu	Gly	Gln			
				165					170					175				
Arg	Gly	Ser	Ser	Arg	Gln	Gly	Ser	Pro	Gly	Ser	Val	Ser	Lys	Gln	Lys			
			180					185						190				
Pro	Cys	Asp	Leu	Pro	Leu	Arg	Leu	Leu	Val	Pro	Thr	Gln	Phe	Val	Gly			
			195				200					205						
Ala	Ile	Ile	Gly	Lys	Glu	Gly	Ala	Thr	Ile	Arg	Asn	Ile	Thr	Lys	Gln			
			210			215						220						
Thr	Gln	Ser	Lys	Ile	Asp	Val	His	Arg	Lys	Glu	Asn	Ala	Gly	Ala	Ala			
				230						235					240			
Glu	Lys	Ser	Ile	Thr	Ile	Leu	Ser	Thr	Pro	Glu	Gly	Thr	Ser	Ala	Ala			
				245					250						255			
Cys	Lys	Ser	Ile	Leu	Glu	Ile	Met	His	Lys	Glu	Ala	Gln	Asp	Ile	Lys			
			260					265					270					
Phe	Thr	Glu	Glu	Ile	Pro	Leu	Lys	Ile	Leu	Ala	His	Asn	Asn	Phe	Val			
			275				280					285						
Gly	Arg	Leu	Ile	Gly	Lys	Glu	Gly	Arg	Asn	Leu	Lys	Lys	Ile	Glu	Gln			
			290			295					300							
Asp	Thr	Asp	Thr	Lys	Ile	Thr	Ile	Ser	Pro	Leu	Gln	Glu	Leu	Thr	Leu			
				310						315					320			
Tyr	Asn	Pro	Glu	Arg	Thr	Ile	Thr	Val	Lys	Gly	Asn	Val	Glu	Thr	Cys			
				325					330						335			
Ala	Lys	Ala	Glu	Glu	Glu	Ile	Met	Lys	Lys	Ile	Arg	Glu	Ser	Tyr	Glu			
			340					345					350					
Asn	Asp	Ile	Ala	Ser	Met	Asn	Leu	Gln	Ala	His	Leu	Ile	Pro	Gly	Leu			
			355				360					365						
Asn	Leu	Asn	Ala	Leu	Gly	Leu	Phe	Pro	Pro	Thr	Ser	Gly	Met	Pro	Pro			
			370			375					380							
Pro	Thr	Ser	Gly	Pro	Pro	Ser	Ala											

			500					505					510				
Leu	Gln	Asn	Leu	Ser	Ser	Ala	Glu	Val	Val	Val	Pro	Arg	Asp	Gln	Thr		
		515						520					525				
Pro	Asp	Glu	Asn	Asp	Gln	Val	Val	Val	Lys	Ile	Thr	Gly	His	Phe	Tyr		
		530				535						540					
Ala	Cys	Gln	Val	Ala	Gln	Arg	Lys	Ile	Gln	Glu	Ile	Leu	Thr	Gln	Val		
545					550					555					560		
Lys	Gln	His	Gln	Gln	Gln	Lys	Ala	Leu	Gln	Ser	Gly	Pro	Pro	Gln	Ser		
			565					570						575			

Arg Arg Lys

<210> 33  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 33  
 catggactgg ctttctgggtt g

21

<210> 34  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens

<400> 34  
 ctgagaaaag ctctggcctt aaac

24